

AD \_\_\_\_\_

Award Number: DAMD17-99-1-9515

TITLE: Determining Gene Expression, Substrate Specificity and  
Natural Substrates for Prostate Associated Proteases

PRINCIPAL INVESTIGATOR: Toshihiko Takeuchi, Ph.D.

CONTRACTING ORGANIZATION: University of California  
San Francisco, California 94143-0962

REPORT DATE: July 2000

TYPE OF REPORT: Annual Summary

PREPARED FOR: U.S. Army Medical Research and Materiel Command  
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;  
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 074-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503				
1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE July 2000	3. REPORT TYPE AND DATES COVERED Annual Summary (1 Jul 99 - 30 Jun 00)		
4. TITLE AND SUBTITLE Determining Gene Expression, Substrate Specificity and Natural Substrates for Prostate Associated Proteases		5. FUNDING NUMBERS DAMD17-99-1-9515		
6. AUTHOR(S) Toshihiko Takeuchi, Ph.D.				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) University of California San Francisco, California 94143-0962  E-MAIL: toshi@foxtrot.ucsf.edu		8. PERFORMING ORGANIZATION REPORT NUMBER		
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)  U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012		10. SPONSORING / MONITORING AGENCY REPORT NUMBER		
11. SUPPLEMENTARY NOTES				
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution unlimited			12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 Words) We work on serine proteases of the chymotrypsin fold due to the wealth of information that exists on structure-function relationships regarding this class of enzymes and the existence of potent and specific inhibitors that are readily available for their inhibition. In addition, understanding of the function of these proteases may lead to further insight into cancer biology and may lead to possible therapeutics. We have focused upon an enzyme known as membrane-type serine protease 1 (MT-SP1) that has been implicated in prostate cancer. During our first funding year, we have determined the in vitro cleavage substrate specificity of the enzyme using positional-scanning combinatorial libraries and substrate phage display. The preferred cleavage sequences were found to be {P4-(Arg/Lys) P3-(X) P2-(Ser) P1-(Arg) P1'-(Ala)} and {P4-(X) P3-(Arg/Lys) P2-(Ser) P1(Arg) P1'(Ala)} where X is a non-basic amino acid. We have also identified two macromolecular substrates, single-chain urokinase-type plasminogen activator (sc-uPA) and protease activated receptor 2 (PAR2). The affinity of MT-SP1 for these key extracellular targets suggests an important role in pathological and regulatory events such as tumor cell invasion and metastasis.				
14. SUBJECT TERMS Prostate Cancer			15. NUMBER OF PAGES 38	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT Unlimited	

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)  
Prescribed by ANSI Std. Z39-18  
298-102

20010228 098

## FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

¶ Where copyrighted material is quoted, permission has been obtained to use such material.

¶ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

¶ Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.


N/A In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and use of Laboratory Animals of the Institute of Laboratory Resources, national Research Council (NIH Publication No. 86-23, Revised 1985).

N/A For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

N/A In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

N/A In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

N/A In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

 6/23/00  
\_\_\_\_\_  
PI - Signature Date

## Table of Contents

Cover.....	1
SF 298.....	2
Foreword.....	3
Table of Contents.....	4
Introduction.....	5
Body.....	5-8
Key Research Accomplishments.....	9
Reportable Outcomes.....	9
Conclusions.....	10
References.....	10
Appendices.....	paper attached

## INTRODUCTION

Numerous studies have shown that proteolytic enzymes play an essential role in tumor invasion, metastasis and angiogenesis (1,2, 3). Proteases may enhance these processes through the release and/or activation of growth factors and the degradation of extracellular matrix. Thus, inhibition of these protease-mediated processes may be a therapeutic strategy for the treatment of metastatic cancer. Indeed, proteases, in general, have proven to be excellent chemotherapeutic targets as evidenced by the HIV protease inhibitors and the large number of successful clinical trials involving protease inhibitors (4). This proposal focuses on serine proteases of the chymotrypsin fold due to the wealth of information that exists on structure-function relationships regarding this class of enzymes and the existence of potent and specific inhibitors that are readily available for their inhibition. In addition, understanding of the function of these proteases may lead to further insight into cancer biology and may lead to possible therapeutics. In particular, we proposed to understand serine protease function by identifying serine protease substrates and substrate specificity. These techniques used and developed should become increasingly useful as the Human Genome Project continues to provide numerous serine protease genes whose products have no known function. These studies should be generalizable to other classes of proteases and to other enzymes such as kinases. In addition, perhaps a greater understanding of prostate and prostate cancer biology can be gained by examining the interaction of the proteases and downstream substrates within a tumor tissue and during various stages of cancer.

## BODY

We recently reported the cloning and initial characterization of membrane-type serine protease 1 (MT-SP1) from the PC-3 human prostatic cancer cell line (5). We have found subcutaneous coinjection of PC-3 cells with wild-type ecotin or ecotin M84R/M85R led to a decrease in the primary tumor size compared to animals in whom PC-3 cells and saline were injected (Melnyk et al, unpublished results). Since both wild-type ecotin and ecotin M84R/M85R are potent, subnanomolar inhibitors of MT-SP1, these studies raise the possibility that MT-SP1 plays an important role in progression of epithelial cancers expressing this protease. Northern blotting showed that MT-SP1 was strongly expressed in the gastrointestinal tract and the prostate, while lower expression levels were observed in the kidney, liver, lung, and spleen. The function of MT-SP1 and its possible role in pathological states is still undetermined. However, potent macromolecular inhibitors of MT-SP1 have been identified and reagent quantities of a

His-tagged fusion of the MT-SP1 protease domain were expressed in *E. coli*, purified and autoactivated (5). Biochemical characterization of the catalytic domain of MT-SP1 may provide insight regarding its physiological role. Therefore, we proposed the following:

**Task 1.** Determine *in vitro* substrate specificity for MT-SP1 and MT-SP2 using a positional scanning-synthetic combinatorial library and substrate phage display (months 1-9).

- A. Assay *in vitro* substrate specificity for MT-SP1 and MT-SP2 using positional scanning library
- B. Synthesize and purify small molecule synthetic substrates for MT-SP1 and MT-SP2, and characterize protease activity against substrates.
- C. Design, create, and screen substrate phage display library.

The *in vitro* characterization of MT-SP2 was problematic due to poor expression levels of the protease. However, we are currently expressing MT-SP2 in a baculovirus expression system, and therefore hope to continue with steps A., B., and C. with MT-SP2. However, we have successfully completed Task 1 with MT-SP1.

**Task 1A:** When a PS-SCL library with the general structure Ac-X-X-X-Lys-AMC was used to profile MT-SP1, the specificity was found to be: (P4 = K>R; P3 = R/K/Q; P2 = S>F/G) (Figure 3). Thus the basic residues lysine and arginine are preferred residues at the P4 position, while these basic residues as well as glutamine are preferred at the P3 position. Interestingly, glycine, serine, and phenylalanine are all well tolerated at P2, despite their difference in size and hydrophobicity. The preference for phenylalanine at the P2 position is not a result of a biased library, since this library has been used to profile other enzymes such as thrombin, and phenylalanine was not cleaved efficiently at this position. This affinity for phenylalanine in the P2 position was also validated using macromolecular substrates, described later.

**Task 1B:** We have synthesized KRSR-AMC, RQSR-AMC, RRSR-AMC and compared these substrates to nTPK-AMC and nTPR-AMC. Consistent with the PS-SCL results, we found that RQSR-AMC had the highest  $k_{cat}$  of 3.5/s, while the nTPR-AMC, lacking the proper specificity has a  $k_{cat}$  of 0.6/s. We continued to Task 1C to further characterize substrate specificity.

**Task 1C:** Two substrate phage libraries were utilized to determine the substrate specificity C-terminal to the scissile bond and to determine whether there are any

interdependencies among the enzyme subsites. The first phage display library was an unbiased library in which P1 was fixed as Arg, while P4-P2 and P1' were completely randomized. The overall cleavage affinities for a given subsite match closely with the results from the PS-SCL. In this phage display library, basic residues appear in P4, although it is not to the same extent observed in the PS-SCL. Similarly glycine is observed at P2, whereas serine was the most favorable residue in PS-SCL. This affinity for glycine at P4 and P2 may be a result of increased flexibility of the peptide resulting in an increased kinetic rate of cleavage for substrate phage. Similar results were observed when substrate phage display was performed on both tissue-type plasminogen activator and uPA (6).

One intriguing finding from the substrate phage display was the apparent dependency between P4 and P3. If P3 is basic, then P4 tends to be non-basic (15 of 17 clones). Similarly, if P4 is basic, then P3 tends to be non-basic (13 of 15 clones). Thus on average, basic residues are most abundant at both P3 and P4, but for each individual cleavage sequence, either P3 or P4 is basic, but not both together. A second substrate phage display library was constructed to explore this possibility. The library was designed based upon the consensus sequence obtained from the PS-SCL. The P3 and P2 position was fixed as a mixture of (R/K/Q) and (G/A/S) respectively, while P4 was allowed to vary. Based upon the observations from the unbiased library, the expectation would be that if P3 is basic, then P4 should be predominantly occupied by neutral side chains. The results from this biased library are displayed in Table Ib, Figure 4b. Indeed, since P3 is constrained to be a basic residue or glutamine, the predominant occupation of P4 is with a neutral residue, further verifying the observed dependency between P4 and P3.

**Task 2.** Generate a human prostate cDNA library for use in the identification of natural substrates for MT-SP1 and MT-SP2 using small pool cDNA expression cleavage screening (months 10-17).

- A. Generate and characterize human prostate cDNA library.
- B. Screen cDNA library using expression cleavage screening.
- C. Isolate and characterize clones derived from screen.

Since Task 1 was completed with MT-SP1, the next phase of the project is to identify possible natural substrates. We initially proposed to use small pool cDNA expression cleavage screening in Task 2; however, there was enough information from the substrate specificity data derived from Task 1 to choose logical candidate substrates for MT-SP1 without using the small pool cDNA expression cleavage screening.

Therefore, we successfully used a candidate approach for the identification of macromolecular substrates of MT-SP1.

Determination of the substrate specificity of MT-SP1 may provide insight into the natural function of the enzyme. Information regarding the peptide substrate specificity combined with the knowledge of enzyme localization led to the testing of logical macromolecular substrates of MT-SP1 that are localized to the extracellular surface. Potential candidates for MT-SP1 cleavage are the protease activated receptors (PAR). PARs are activated by cleavage of a single site in their N-terminal exodomains. Of the four PARs known, only PAR2's cleavage site contains a basic residue in P4 (Ser) or P3 (Lys) and a small residue or phenylalanine in P2 (Gly). This led to the expectation that MT-SP1 would activate PAR2, but not activate PAR1, 3, and 4. The activation of these receptors was tested by injecting xenopus oocytes with PAR cRNAs and monitoring activation of the receptors upon addition of exogenous protease. Addition of MT-SP1 catalytic domain to a final concentration of 1, 10 or 100 nM led to activation of PAR2 at 10 and 100 nM, while activation of PAR1, 3 or 4 was not observed at any of the three concentrations. This specificity for PAR2 was seen even when PAR2 was expressed at much lower levels than PAR1, PAR3, or PAR4. In addition to MT-SP1, 0.5 nM trypsin was also shown to activate PAR2. The PAR1, 3, and 4 receptors were functional, since these receptors were activated by thrombin. This data shows that the MT-SP1 catalytic domain can selectively activate PAR2 over the other receptors, validating the substrate specificity determined by PS-SCL and phage display.

Another potential substrate that is consistent with the MT-SP1 specificity profile is single-chain urokinase-type plasminogen activator (sc-uPA). Sc-uPA contains a neutral P4 (Pro) with a basic P3 (Arg), a Phe at P2, and a Lys at P1. The macromolecular substrate sc-uPA is a good test for the expected P2 Phe specificity. Indeed, sc-uPA is an excellent substrate for MT-SP1. There is an increase in proteolytic activity that is dependent upon the presence of both sc-uPA and MT-SP1 that increases in time. Concomitant with this increase in activity is cleavage of sc-uPA to the A and B-chain components, as expected upon activation of the enzyme. Under the same conditions, plasminogen was not activated by MT-SP1. Plasminogen has a small P2 residue (Gly), but lacks a basic P4 or P3 residue. These results further verify the specificity of MT-SP1 derived from the substrate libraries.

**Task 3** was proposed for months 18-24, and should be carried out as planned previously in the second year of research.

## Key Research Accomplishments

- Assayed *in vitro* substrate specificity for MT-SP1 using positional scanning library
- Synthesized and purified small molecule synthetic substrates for MT-SP1 and MT-SP2, and characterized protease activity against substrates.
- Designed, created, and screened two substrate phage display libraries against MT-SP1
- The preferred cleavage sequences for MT-SP1 were found to be {P4-(Arg/Lys) P3-(X) P2-(Ser) P1-(Arg) P1'-(Ala)} and {P4-(X) P3-(Arg/Lys) P2-(Ser) P1(Arg) P1'(Ala)} where X is a non-basic amino acid.
- Using the preferred substrate cleavages, macromolecular substrates for MT-SP1 were identified:
  - 10 nM soluble MT-SP1 protein activates protease activated receptor-2 (PAR2) but not PAR1, PAR3, or PAR4.
  - 1 nM soluble MT-SP1 protein activates single-chain urokinase type plasminogen activator, but not plasminogen.
- The membrane localization of MT-SP1 and its affinity for these key extracellular substrates suggests a role of the proteolytic activity in regulatory events.

## Reportable Outcomes

Paper:

T. Takeuchi, J. L. Harris, W. Huang, K. W. Yan, S. R. Coughlin, and C. S. Craik "Cellular localization of membrane-type serine protease 1 and identification of protease activated receptor-2 and single-chain urokinase-type plasminogen activator as substrates", *J. Biol. Chem.*, in press, published online at [www.jbc.com](http://www.jbc.com)

Talk:

T. Takeuchi, J. L. Harris, F. Elfman, W. Huang, K. Yan, S. R. Coughlin, M. A. Shuman, and C. S. Craik "Using Macromolecular Protease Inhibitors to Dissect Complex Biological Processes: Identification and Characterization of a Membrane-Type Serine Protease in Epithelial Cancer and Normal Tissue", International Symposium on Proteases: Basic Aspects and Clinical Relevance, Montebello, Quebec, 2000.

## Conclusions

Significant progress has been made toward the tasks outlined in the initial proposal. The goal of tasks 1 and 2 (months 1-17) were to determine the substrate specificity of MT-SP1 and MT-SP2 and identify possible macromolecular substrates of these enzymes. These goals have been completed for MT-SP1, and we are confident that expression of MT-SP2 should be possible using the baculovirus system. Furthermore, the synthetic combinatorial libraries and the phage display libraries have already been created. Therefore, the characterization of MT-SP2 should be relatively straightforward, since much of the time goes into the preparation of these libraries. We are confident that we can perform the proposed studies and are excited about the research in this upcoming year.

## References

---

1. Aznavoorian, S., Murphy, A. N., Stetler-Stevenson, W. G., and Liotta, L. A. (1993) "Molecular aspects of tumor cell invasion and metastasis." *Cancer* **71**, 1368-1383.
2. Chen, W. T. (1992) "Membrane proteases: roles in tissue remodeling and tumor invasion." *Curr. Opin. Cell Biol.* **4**, 802-809.
3. Liotta, L. A., Steeg, P. S., and Stetler-Stevenson, W. G. (1991) "Cancer metastasis and angiogenesis: an imbalance of positive and negative regulation." *Cell* **64**, 327-336.
4. Deeks S. G., Smith M., Holodniy M., Kahn, J. O. (1997) HIV-1 protease inhibitors. A review for clinicians. *Jama*, **277**, 145-53
5. Takeuchi, T., Shuman, M. A., and Craik, C. S. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 11054-11061
6. Ke, S. -H., Coombs, G. S., Tachias, K., Navre, M., Corey, D. R., and Madison, E. L. (1997) *J. Biol. Chem.* **272**, 16603-16609

Cellular localization of membrane-type serine protease 1 and identification of protease activated receptor-2 and single-chain urokinase-type plasminogen activator as substrates

Toshihiko Takeuchi\*, Jennifer L. Harris\*, Wei Huang<sup>§</sup>, Kelly W. Yan\*\*, Shaun R. Coughlin<sup>§</sup>, and Charles S. Craik\*<sup>◇</sup>

\*Department of Pharmaceutical Chemistry and Biochemistry and Biophysics,

<sup>§</sup>Cardiovascular Research Institute, University of California, San Francisco, CA 94143

\*\*Center for Biomedical Laboratory Science, San Francisco State University, San Francisco, Ca 94132

<sup>◇</sup>Corresponding author:

Fax: 415-502-8298, Phone: 415-476-8146, Email: [craik@cgl.ucsf.edu](mailto:craik@cgl.ucsf.edu)

Running title: Characterization of MT-SP1 and identification of substrates

Abbreviations:

MT-SP1, membrane-type serine protease 1; CUB, complement factor 1R-urchin embryonic growth factor-bone morphogenetic protein; LDL, low density lipoprotein; sc-uPA, single-chain urokinase-type plasminogen activator; pNA, *p*-nitro-anilide; PAR, protease activated receptor; PS-SCL, positional scanning-synthetic combinatorial library.

## Summary

Membrane-type serine protease 1 (MT-SP1) was recently cloned, and we now report its biochemical characterization. MT-SP1 is predicted to be a type-II transmembrane protein with an extracellular protease domain. This localization was experimentally verified using immunofluorescent microscopy and a cell surface biotinylation technique. The substrate specificity of MT-SP1 was determined using a positional scanning-synthetic combinatorial library and substrate phage techniques. The preferred cleavage sequences were found to be {P4-(Arg/Lys) P3-(X) P2-(Ser) P1-(Arg) P1'-(Ala)} and {P4-(X) P3-(Arg/Lys) P2-(Ser) P1(Arg) P1'(Ala)} where X is a non-basic amino acid. Protease activated receptor 2 (PAR2) and single-chain urokinase-type plasminogen activator (sc-uPA) are proteins that are localized to the extracellular surface and contain the preferred MT-SP1 cleavage sequence. The ability of MT-SP1 to activate PARs was assessed by exposing PAR-expressing *Xenopus* oocytes to the soluble MT-SP1 protease domain. The latter triggered calcium signaling in PAR2-expressing oocytes at 10 nM but failed to trigger calcium signaling in oocytes expressing PAR1, PAR3, or PAR4 at 100nM. sc-uPA was activated using catalytic amounts of MT-SP1 (1 nM), but plasminogen was not cleaved under similar conditions. The membrane localization of MT-SP1 and its affinity for these key extracellular substrates suggests a role of the proteolytic activity in regulatory events.

## Introduction

We recently reported the cloning and initial characterization of membrane-type serine protease 1 (MT-SP1) from the PC-3 human prostatic cancer cell line (1). Northern blotting showed that MT-SP1 was strongly expressed in the gastrointestinal tract and the prostate, while lower expression levels were observed in the kidney, liver, lung, and spleen. The function of MT-SP1 and its possible role in pathological states is still undetermined. However, potent macromolecular inhibitors of MT-SP1 have been identified and reagent quantities of a His-tagged fusion of the MT-SP1 protease domain were expressed in *E. coli*, purified and autoactivated (1). Biochemical characterization of the catalytic domain of MT-SP1 may provide insight regarding its physiological role.

MT-SP1 is predicted to be a modular, type-II transmembrane protein that contains a signal/anchor domain, two CUB domains, four LDL receptor repeats and a serine protease domain (1). The mouse homolog of MT-SP1, called epithin, recently was reported to be strongly expressed in fetal thymic stromal cells and highly expressed in a thymic epithelial nurse cell line (2). Another report describes the N-terminal sequencing of a protein called matriptase from human breast milk (3). The reported matriptase sequence is included in the translated sequence for the cDNA of MT-SP1. The matriptase cDNA reported appears to be a partial MT-SP1 cDNA, lacking 516 of the coding nucleotides. However, since the matriptase cDNA encodes a possible initiating methionine, alternative splicing could yield a protein lacking the N-terminal region of MT-SP1.

While MT-SP1 and epithin are predicted to be type-II transmembrane proteins, the reported matriptase cDNA lacks the 5' end of the MT-SP1 cDNA and therefore the translated sequence lacks the signal/anchor domain, leading to a predicted secreted protein. Determining the cellular localization of the protein could help resolve this discrepancy and may also provide clues for understanding the function of the protein. For example, another structurally similar membrane-type serine protease, enteropeptidase, is involved in a proteolytic cascade by which activation of trypsinogen leads to activation of other digestive

proteases (4). The membrane localization is essential to restrict the activation of trypsinogen to the enterocytes of the proximal small intestine. Since MT-SP1 is also predicted to be a membrane-type protease, localization to the membrane may be essential to the proper function of the enzyme. The localization of MT-SP1 is addressed in this work using immunofluorescent localization, immunoblot analysis and cell-surface biotinylation experiments.

Further understanding of the role of MT-SP1 may be obtained by characterizing the activity of the protease domain. Reagent quantities of a His-tagged fusion of the MT-SP1 protease domain were expressed in *E. coli*, purified and autoactivated, allowing determination of MT-SP1 substrate specificity. Synthetic substrates are typically used to determine the specificity of proteases. However, the use of single substrates can be tedious for synthetic peptide substrates that contain multiple amino acid residues; the exhaustive analysis of each substrate for all combinations of amino acids at multiple positions rapidly becomes impractical. By using pools of substrates through combinatorial techniques, rapid determination of the full specificity profile for an enzyme can be obtained. Two methods have been employed to determine the substrate specificity of the MT-SP1 protease domain: positional-scanning synthetic combinatorial libraries (PS-SCL) (5,6,7,8) and substrate phage display (8,9,10).

PS-SCL of fluorogenic peptide substrates has been a very powerful tool for determining protease specificity for proteases that require an Asp in the P1<sup>1</sup> (11) position (6-8). However, the synthetic strategy used to make the P1-Asp library is not generalizable to all amino acids. However, a strategy allowing diversity at P1 has been achieved through nucleophilic displacement of the peptide library from the solid support by condensation with a fluorogenic 7-amino-4-methylcoumarin (AMC)-derivatized amino acid (12). This

---

<sup>1</sup> Nomenclature for the substrate amino acid preference is P<sub>n</sub>, P<sub>n-1</sub>,...,P<sub>2</sub>, P<sub>1</sub>, P<sub>1</sub>', P<sub>2</sub>',...,P<sub>m-1</sub>',P<sub>m</sub>'. Amide bond hydrolysis occurs between P<sub>1</sub> and P<sub>1</sub>'. S<sub>n</sub>, S<sub>n-1</sub>,..., S<sub>2</sub>, S<sub>1</sub>, S<sub>1</sub>', S<sub>2</sub>', ..., S<sub>m-1</sub>', S<sub>m</sub>' denotes the corresponding enzyme binding sites (Schechter, I. and Berger, A. (1967) *Biochem. Biophys. Res. Commun.* **27**, 157-162).

strategy was used to create a PS-SCL library with the general structure Ac-X-X-X-Lys-AMC (12), and can be applied to enzymes such as MT-SP1 that have basic P1 specificity. Since PS-SCL cannot be used to determine the specificity C-terminal to the scissile bond (prime side: P1', P2',...,Pn') due to the requirement of the AMC in the P1' position, substrate phage display was utilized (8,9,10). An inexpensive, accessible phage display technique utilizes a cleavable peptide sequence that is inserted between a histidine tag affinity anchor and the M13 phage coat protein, pIII. Bacteriophage containing preferred peptide recognition sequences for a given protease are cleaved from the resin, recovered and amplified, while uncleaved phage remain bound to the Ni(II) resin. After several rounds of cleavage and subsequent amplification of the phage, the phagemid DNA plasmids can be sequenced and analyzed for protease substrate specificity preferences (8).

Together, these techniques allowed the determination of the extended substrate specificity of MT-SP1; this specificity was used to identify protease-activated receptor2 (PAR2) and single-chain urokinase-type plasminogen activator (sc-uPA) as macromolecular substrates of MT-SP1. PAR2 is expressed in vascular endothelial cells and in a variety of epithelial cells and may function in inflammation, cytoprotection, and/or cell adhesion (13, 14, 15, 16), while uPA has been implicated in tumor cell invasion and metastasis (17, 18). Therefore, this study raises potential biological and pathological consequences of MT-SP1 activity.

## **Experimental Procedures**

### *Materials*

All primers used were synthesized on an Applied Biosystems 391 DNA synthesizer. All restriction enzymes were purchased from New England Biolabs (Beverly, MA). Automated DNA sequencing was carried out on an Applied Biosystems 377 Prism sequencer, and manual, chain termination, DNA sequencing was carried out under standard conditions. Deglycosylation was performed using PNGase F (New England Biolabs; Beverly, MA). All other reagents were of the highest quality available and purchased from Sigma (St. Louis, Mo) or Fisher (Pittsburgh, Pa) unless otherwise noted.

### *Antibody production and immunoblot analysis*

Polyclonal antiserum against purified His-MT-SP1 protease domain was raised in rabbits (Covance Corporation; Richmond, CA). This antiserum was further purified by binding and elution from an antigen column, which had the His-tag fusion of the inactive Ser805Ala MT-SP1 protease domain covalently linked to the column using NHS-activated Sepharose 4 Fast Flow (Pharmacia Amersham; Piscataway, NJ). Immunoblot analysis was performed as described previously (19). Antibody-bound protein bands were detected using a goat anti-rabbit horseradish peroxidase-conjugated secondary antibody (Pierce; Rockford, IL) and enhanced chemiluminescence (Pharmacia Amersham; Piscataway, NJ).

### *Cell culture and immunofluorescence*

The PC-3 (CRL-1435) and HeLa S3 (CCL-2.2) cell lines were purchased from ATCC (Manassas, VA) and grown according to the instructions provided by ATCC. The cells were plated on glass cover slips and were stained as described previously (20). Permeabilization was performed with 0.5 % Triton X-100 in phosphate buffered saline (PBS). Non-permeabilized cells were treated with PBS. The primary antibody was either affinity purified anti-MTSP1 at 1:500 dilution, monoclonal anti-uPAR antibody at 1:500 dilution, or an anti-vimentin antibody at 1:100 dilution in 5% goat serum. The appropriate biotinylated anti-mouse or anti-rabbit secondary antibody (Jackson ImmunoResearch; West

Grove, PA) was used at 1:1000 dilution in 5% goat serum. Subsequently, the cells were treated with FITC-conjugated streptavidin at 1:500 dilution in 5% goat serum. Cover slips were mounted using Gel Mount (Biomed; Foster City, CA) and samples were imaged using an Olympus BX60 fluorescent microscope. Each of the antibody incubation steps were for an hour with three five minute PBS washes in between each antibody incubation. Cell lysates were prepared using PBS containing 1% Triton X-100 and 5 mM ethylenediamine tetraacetic acid (EDTA).

#### *Cell Surface Biotinylation*

The cell impermeable sulfo-NHS-biotin (sulfosuccinimidobiotin) (Pierce, Rockford, IL) was used to biotinylate surface proteins as described earlier (21). Greater than 95% of the cells remained impermeable as assayed by the impermeability of the cells to trypan blue. Cell lysates were prepared using PBS containing 1% Triton X-100 and 5 mM EDTA. Biotinylated proteins were captured using streptavidin-agarose (Life Technologies, Gaithersburg, MD), and electrophoresed on a 10% SDS polyacrylamide gel. Immunoblot analysis was performed as described above. No MT-SP1 was observed in the non-biotinylated PC-3 extracts.

#### *Creation of a P1-Lysine Positional Scanning Combinatorial Library*

The detailed synthesis and characterization of the combinatorial library used in this study are described elsewhere (12). Three support-bound sub-libraries were prepared (P2, P3, P4) employing an alkanesulfonamide linker (22) and solid-phase peptide synthesis. Each sub-library consisted of 19 resins (one unnatural amino acid, norleucine, was included, while cysteine and methionine were excluded) for which a single position was spatially addressed by the coupling of a single amino acid. The two remaining positions of each resin were supplied by the coupling of isokinetic mixtures of amino acid derivatives (23) to give a resin-bound mixture of 361 different peptides. The 57 resins comprising the entire PS-SCL were put into individual wells and cleaved from the resin with a lysine-coumarin derivative. Filtration, side-chain deprotection, and concentration provided a PS-SCL of 57 wells

containing 361 tetrapeptide-coumarin derivatives per well for a total of 6,859 peptide substrates per library.

#### *Enzymatic Assay of the PS-SCL*

The concentration of MT-SP1 was determined by active-site titration as described earlier (1). Substrates from the PS-SCL were dissolved in DMSO. Approximately  $2.5 \times 10^{-9}$  mol of each sub-library (361 compounds) were added to 57 wells of a 96-well Microfluor White "U" bottom plate (Dynex Technologies, Chantilly, VA). Final substrate concentration was approximately 0.25  $\mu$ M, making the hydrolysis of the AMC group directly proportional to the specificity constant,  $k_{\text{cat}}/K_m$ . Hydrolysis reactions were initiated by the addition of enzyme (1 nM) and monitored fluorometrically with a Perkin Elmer LS50B Luminescence Spectrometer 96-well plate reader, with excitation at 380 nm and emission at 460nm. Assays were performed in a buffer containing 50 mM Tris, pH 8.8, 100 mM NaCl, 1% DMSO (from substrates) and 0.01% Tween-20.

#### *Creation of His-tagged Substrate Phage Libraries*

The phagemid pHisX3P3, derived from pBS was used as described previously (8). In the biased library, the cleavage sequence was based upon the consensus sequence derived from the PS-SCL library results: (X-(R/Q/K)-(S/A/G)-R-XX), where X can encode any amino acid in the P4 position, P3 encodes at least arginine, glutamine, and lysine, P2 encodes at least serine, alanine, and glycine, P1 is fixed as arginine, and both P1' and P2' are completely randomized. In the unbiased library, the randomized peptide sequence encoded in the vector is XXXRX, where X can encode any amino acid. In this cleavage sequence, P1 is fixed as arginine while P4-P2 and P1' are randomized. The degenerate oligonucleotides synthesized to create the library contained the following randomized sequences (where N indicates equimolar concentrations of A, C, G and T; S indicates equimolar concentrations of G and C, M indicates equimolar mixtures of A and C, R indicates equimolar mixtures of A and G, and K indicates equimolar mixtures of G and T.): NNS MRG KSS AGG NNS NNS (biased library) and NNS NNS NNS AGA (unbiased library). The phage library and

phagemid vector were constructed by ligation into the cut pHisX3P3 vector followed by electroporation of the ligated vector into XL2-Blue MRF<sup>+</sup> cells (Stratagene, La Jolla, CA) as described earlier (8). In the biased library, the transformation efficiency was  $2.4 \times 10^7$  individual clones, and the transformation efficiency of the unbiased library was  $2.1 \times 10^7$  individual clones, allowing for >99% completeness of each library.

#### *His-tagged Substrate Phage Cleavage*

Two hundred microliters of nickel(II)-nitrilotriacetic acid resin (Qiagen, Santa Clarita, CA) was washed with 10 mL of activity buffer (50 mM Tris, pH 8.8, 100 mM NaCl, 0.1% Tween 20). Phage particles ( $10^9$ ) were added to the washed Ni(II) resin and bound with gentle agitation for 1 hour. The Ni(II) resin subsequently was washed with 5 mL of activity buffer and gently agitated for 30 minutes. This washing step was repeated for a total of four 30 minute washes. The activity buffer was removed, and the bound phage were eluted twice with 0.5 mL of activity buffer containing 0.5 M imidazole. The imidazole was removed using a PD-10 column (Amersham Pharmacia; Piscataway, NJ). The resulting solution was concentrated to a volume of 0.5 mL using a centricon-100 filter concentrator (Millipore; Bedford, MA). The phage then were cleaved with 1 nM recombinant MT-SP1 protease domain for 1 hour at 37° C. A control sample lacking MT-SP1 was used to monitor binding of uncleaved phage to the Ni(II) resin. The cleaved phage were added to 200  $\mu$ L of washed Ni(II) resin to rebind uncut phage and allowed to bind for 3 hours. Phage that are cleaved by MT-SP1 lack the His-tag and will not bind to Ni(II) resin. These unbound phage were eluted and amplified as described earlier (8) and the cleavage round was repeated. Five rounds of panning were completed with the biased library before sequencing, while eight rounds of panning were performed for the unbiased library.

#### *Molecular modeling of the MT-SP1-substrate complex*

All modeling was performed using the Biopolymer and Homology modules within Insight II (Molecular Simulations; San Diego, Ca). The MT-SP1 amino acids were threaded onto the  $\beta$ -tryptase crystal structure (24) (1AOL). A model of an inhibitor bound MT-SP1

structure was produced by using a trypsin-ecotin crystal structure (25). The trypsin-ecotin crystal structure was modeled onto the MT-SP1 structure by overlaying the trypsin and MT-SP1 protease domains; subsequently the trypsin was removed from the model. The active-site protease binding loop of ecotin was used as a model of a substrate binding to the MT-SP1 active site. The preferred side chain rotamers of the modeled substrate were explored manually to maximize interaction with the MT-SP1 active site.

#### *Assay for PAR activation*

cDNAs encoding hPAR1, mPAR2, hPAR3 and hPAR4 tagged with a FLAG epitope were used (26, 27, 28, 29). *Xenopus* oocytes were microinjected with 25 ng hPAR1, 0.25 ng mPAR2, 25 ng hPAR3 and 2 ng hPAR4 cRNA per oocyte. <sup>45</sup>Ca release triggered by soluble MT-SP1 protease domain was measured (26). PAR expression on the oocyte surface was quantitated using a colorimetric assay that measures the level of the FLAG tag, which was displayed at the extracellular N-terminus of each PAR (26, 30).

#### *Cleavage of sc-uPA*

sc-uPA (5  $\mu$ M) was incubated with MT-SP1 (1 nM) in 50 mM Tris, pH 8.8, 100 mM NaCl at 37 °C. At specified intervals, an aliquot was withdrawn and split into two portions. The first portion was assayed for activity against Spectrozyme UK (carbobenzoxycarbonyl-L-glutamyl( $\alpha$ -t-butoxy)-glycyl-arginine-p-nitroanilide; American Diagnostica; Greenwich, CT), while the second portion was boiled in sample buffer (125 mM Tris-HCl, pH 6.8, 4% SDS, 10% 2-mercaptoethanol, 20 % glycerol) and subjected to immunoblot analysis. Immunoblots were prepared as described above for anti-MT-SP1 immunoblots, except that polyclonal rabbit anti-human uPA antibodies (American Diagnostica; Greenwich, CT) were used as the primary antibody.

## Results

### *Migration pattern in SDS-PAGE and cell surface localization of MT-SP1*

MT-SP1 is predicted to be a 95kD protein, and upon activation of the protease domain, a disulfide link is predicted to tether the catalytic domain to the non-catalytic domains (Figure 1a). Under non-reducing conditions, immunoblotting of Triton extracts derived from PC-3 cells shows a doublet at approximately 80 kD (Figure 1b, lane 4) using polyclonal antibodies directed against the soluble, recombinant MT-SP1 protease domain (Figure 1a3 and 1b, lane 1). Under reducing conditions, the predicted disulfide linkage between the catalytic and non-catalytic domain should be severed, resulting in release of the proteolytic domain. Indeed a band at 87 kD and a band at 29 kD are observed (Figure 1b, lane 3, 5). Upon deglycosylation of the PC-3 Triton extract with PNGase F, only a single band is observed at 85 kD, and the band attributed to the protease domain decreases to the size of the recombinant MT-SP1 (Figure 1b, lane 6). This proposed glycosylation is consistent with the predicted N-linked glycosylation sites in the pro-domain (109, 302, 485) and the protease domain (residue 771) (see reference 1). The band at 85 kD is smaller than the predicted 95 kD of the full-length protein. This decrease in size may be due to significant folding of the protease even under reducing conditions with SDS. This decrease in size appears more pronounced under non-reducing conditions in the presence of SDS (Figure 1b, lane 4), where the protease domain appears to be only 80 kD in size. *In vitro* transcription-translation of full length MT-SP1 cDNA results in a band at 85 kD (Figure 1b, lane 7), which is the same size as the deglycosylated MT-SP1 (Figure 1b, lane 6), supporting the hypothesis that the full-length protein runs smaller than the predicted molecular weight. The absence of the band at 29 kD in this sample is presumably due to the reducing environment in the *in vitro* transcription/translation mixture, which likely prevents proper folding and activation of the protease.

MT-SP1 is predicted to be an integral membrane protein localized to the extracellular surface through a signal/anchor domain (1). The extracellular surface

localization of MT-SP1 was verified using two independent techniques: biotinylation of cell surface proteins and immunofluorescence. Cell surface localization was determined by biotinylating cell surface proteins using a non-permeable biotinylation reagent (21, 31). After removal of unreacted biotin, the cells are lysed with 1% Triton-X-100 and 5 mM EDTA in PBS and biotinylated proteins are bound to streptavidin immobilized agarose. SDS-PAGE followed by immunoblotting with anti-MT-SP1 antibodies showed the presence of MT-SP1 in the biotinylated cell lysate (Figure 1b, Lane 8), while no MT-SP1 was observed in the non-biotinylated lysate (data not shown). The non-permeability of the cells was verified with trypan blue, which showed that >95% of the cells were intact. The extracellular localization of MT-SP1 was independently verified using immunofluorescent microscopy. Figure 2a shows extracellular staining of PC-3 cells under non-permeablizing conditions when treated with rabbit antiserum directed against the MT-SP1 protease domain. Similar staining patterns are observed with treatment against the urokinase plasminogen activator receptor (data not shown). The specificity of anti-MT-SP1 antigen interaction was characterized using both recombinant MT-SP1 (Figure 1b, lane 1), and HeLa S3 cells, which do not express MT-SP1 (Figure 1b, lane 3). Little fluorescence staining is observed for HeLa S3 cells (Figure 2b), suggesting that the observed immunostaining in PC-3 cells is due to specific interaction of the antibodies with MT-SP1 protein. It should be noted that expression in COS cells of MT-SP1 in which the signal/anchor domain was deleted, resulted MT-SP1 that remained bound to the cell surface (data not shown), suggesting that domains other than the signal/anchor are involved in cell surface interactions.

#### *Determination of MT-SP1 substrate specificity*

When a PS-SCL library with the general structure Ac-X-X-X-Lys-AMC (12) was used to profile MT-SP1, the specificity was found to be: (P4 = K>R; P3 = R/K/Q; P2 = S>F/G) (Figure 3). Thus the basic residues lysine and arginine are preferred residues at the P4 position, while these basic residues as well as glutamine are preferred at the P3 position.

Interestingly, glycine, serine, and phenylalanine are all well tolerated at P2, despite their difference in size and hydrophobicity. The preference for phenylalanine at the P2 position is not a result of a biased library, since this library has been used to profile other enzymes such as thrombin, and phenylalanine was not cleaved efficiently at this position (12). This affinity for phenylalanine in the P2 position was also validated using macromolecular substrates, described below.

Two substrate phage libraries were utilized to determine the substrate specificity C-terminal to the scissile bond and to determine whether there are any interdependencies among the enzyme subsites. The first phage display library was an unbiased library in which P1 was fixed as Arg, while P4-P2 and P1' were completely randomized. The results of this library are shown in tabular form in Table Ia, where individual peptide cleavage sequences can be observed. However, the overall cleavage affinities for a given subsite are better displayed in graphical format as shown in Figure 4a. The substrate specificity observed in substrate phage display match closely with the results from the PS-SCL (Figure 3). In this phage display library, basic residues appear in P4, although it is not to the same extent observed in the PS-SCL. Similarly glycine is observed at P2, whereas serine was the most favorable residue in PS-SCL. This affinity for glycine at P4 and P2 may be a result of increased flexibility of the peptide resulting in an increased kinetic rate of cleavage for substrate phage. Similar results were observed when substrate phage display was performed on both tissue-type plasminogen activator and uPA (32).

One intriguing finding from the substrate phage display (Table Ia) was the apparent dependency between P4 and P3. If P3 is basic, then P4 tends to be non-basic (15 of 17 clones). Similarly, if P4 is basic, then P3 tends to be non-basic (13 of 15 clones). Thus on average, basic residues are most abundant at both P3 and P4, but for each individual cleavage sequence, either P3 or P4 is basic, but not both together. A second substrate phage display library was constructed to explore this possibility. The library was designed based upon the consensus sequence obtained from the PS-SCL. The P3 and P2 positions were

fixed as a mixture of (R/K/Q) and (G/A/S) respectively, while P4 was allowed to vary. Based upon the observations from the unbiased library, the expectation would be that if P3 is basic, then P4 should be predominantly occupied by neutral side chains. The results from this biased library are displayed in Table Ib, Figure 4b. Indeed, since P3 is constrained to be a basic residue or glutamine, the predominant occupation of P4 is with a neutral residue, further verifying the observed dependency between P4 and P3.

To assist in defining the molecular determinants of substrate specificity, a homology model of MT-SP1 was constructed. The MT-SP1 amino acids were threaded onto the  $\beta$ -trypsin crystal structure (24) (1AOL). A substrate for MT-SP1 was modeled in the putative binding pocket using the active-site protease binding loop of ecotin that was derived from an ecotin-trypsin crystal structure (25). The S1 subsite specificity for basic amino acids can be attributed to the presence of aspartate 799 at the bottom of this S1 binding pocket. The S2 subsite is predicted to be a shallow groove, however a Phe in the P2 position would be expected to make favorable interactions with Phe706 of MT-SP1. Asp 828 of MT-SP1 could potentially form a salt bridge with either P4 or P3 depending on the conformation of the side chain, resulting in basic specificity in both P4 and P3. The active site from this MT-SP1 model complexed to a substrate Arg-Phe-Gln-Arg in P1 through P4 respectively is displayed in Figure 5.

#### *Macromolecular Substrate Determination*

Determination of the substrate specificity of MT-SP1 may provide insight into the natural function of the enzyme. Information regarding the peptide substrate specificity combined with the knowledge of enzyme localization led to the testing of logical macromolecular substrates of MT-SP1 that are localized to the extracellular surface. Potential candidates for MT-SP1 cleavage are the protease activated receptors (see 33 and references therein). PARs are activated by cleavage of a single site in their N-terminal exodomains. Of the four PARs known, only PAR2's cleavage site contains a basic residue in P4 (Ser) or P3 (Lys) and a small residue or phenylalanine in P2 (Gly). This led to the

expectation that MT-SP1 would activate PAR2, but not activate PAR1, 3, and 4. The activation of these receptors was tested by injecting xenopus oocytes with PAR cRNAs and monitoring activation of the receptors upon addition of exogenous protease. Addition of MT-SP1 catalytic domain to a final concentration of 1, 10 or 100 nM led to activation of PAR2 at 10 and 100 nM, while activation of PAR1, 3 or 4 was not observed at any of the three concentrations (Figure 6a). This specificity for PAR2 was seen even when PAR2 was expressed at much lower levels than PAR1, PAR3, or PAR4 (Figure 6b). In addition to MT-SP1, 0.5 nM trypsin was also shown to activate PAR2. The PAR1, 3, and 4 receptors were functional, since these receptors were activated by thrombin. This data shows that the MT-SP1 catalytic domain can selectively activate PAR2 over the other receptors, validating the substrate specificity determined by PS-SCL and phage display.

Another potential substrate that is consistent with the MT-SP1 specificity profile is single-chain urokinase-type plasminogen activator (sc-uPA). Sc-uPA contains a neutral P4 (Pro) with a basic P3 (Arg), a Phe at P2, and a Lys at P1. The macromolecular substrate sc-uPA is a good test for the expected P2 Phe specificity. Indeed, sc-uPA is an excellent substrate for MT-SP1 as shown in Figure 7. There is an increase in proteolytic activity that is dependent upon the presence of both sc-uPA and MT-SP1 that increases in time (Figure 7). Concomitant with this increase in activity is cleavage of sc-uPA to the A and B-chain components, as expected upon activation of the enzyme (Figure 7). Under the same conditions, plasminogen was not activated by MT-SP1 (data not shown). Plasminogen has a small P2 residue (Gly), but lacks a basic P4 or P3 residue. These results further verify the specificity of MT-SP1 derived from the substrate libraries.

Plasmin has been shown to activate sc-uPA *in vitro* (34); this activation presumably would represent a feedback cycle where a small amount of active uPA would activate plasminogen to plasmin, and the resulting plasmin would activate sc-uPA. Other enzymes have also been reported to activate sc-uPA, including plasma kallikrein (35), cathepsin B (36), cathepsin L (37), mast cell tryptase (38) and prostate-specific antigen (39). In these

studies, sc-uPA was activated using a substrate to enzyme ratio of 30:1, 10:1, 200:1, 50:1 and 10:1 respectively. Under similar conditions, MT-SP1 activates sc-uPA at a substrate:enzyme ratio of 5,000:1. These assays were performed with 1 nM MT-SP1, suggesting highly potent activation of sc-uPA (Figure 7).

## Discussion

We previously reported the cloning and characterization of MT-SP1 derived from the cDNA of the PC-3 human prostatic carcinoma cell line (1). From the translation of the cDNA, MT-SP1 was predicted to be a type-II transmembrane protein. A partial MT-SP1 cDNA had been reported by another laboratory and referred to as "matriptase"(3). The matriptase cDNA lacks the 5' coding region of the MT-SP1 cDNA, resulting in the truncation of the predicted signal anchor domain and instead, is reported to contain a signal peptide. We report immunofluorescence and cell surface biotinylation studies that show MT-SP1 protein is localized to the extracellular cell surface. Moreover, earlier work from the same laboratory that published the matriptase cDNA clone supports the extracellular surface localization (40). In that work, a protein that cross-reacts with matriptase antibodies shows extracellular localization on the surface of breast cancer cells using a cell-surface biotinylation assay and subcellular fractionation further localizes matriptase to the membrane. Their conclusion was that the protein that cross-reacts with matriptase antibodies is an integral membrane protein. Therefore, these data are consistent with the presence of a signal/anchor transmembrane domain in the translated MT-SP1 cDNA, and inconsistent with the presence of a signal peptide as suggested for the matriptase cDNA translation.

One possible explanation for the observed soluble forms of MT-SP1/matriptase protein is through shedding from the extracellular surface. For example, the protein sequenced for the matriptase clone was isolated from breast milk and not from the extracellular surface of cells (3). N-terminal amino acid sequencing showed sequence corresponding to amino acid 350-358 in the MT-SP1 protein translation, and amino acid 228-236 in the matriptase translation, suggesting that the form of MT-SP1 isolated in breast milk most likely is cleaved from the extracellular surface and released into milk. These data, therefore do not conflict with the proposed localization and protein translation for MT-SP1. Another possibility is that the matriptase clone is produced through alternative splicing, resulting in a soluble form of the protein. Isolation and N-terminal sequencing of the

soluble forms may be necessary to differentiate between shed forms of the protein and secreted forms of the protein.

MT-SP1 protein has a predicted molecular weight of 95 kD, while matriptase has a predicted size of 76 kD. The previous matriptase studies reported that the protein isolated from breast cancer cells is 80 kD under non-reducing conditions (40, 41). MT-SP1 under non-reducing conditions has an apparent size of 80 kD (Figure 1, lane 4). However, deglycosylated and reduced MT-SP1 derived from PC-3 cells (Figure 1, lane 6) has an apparent size of 87 kD. Thus, there appears to be significant folding of the protein under non-reducing conditions, leading to a molecular weight that is smaller than the predicted molecular weight. *In vitro* transcription/translated product from the full length MT-SP1 cDNA clone (Figure 1, lane 7) also appears to be 87 kD, therefore the full-length protein may run slightly smaller than the expected 95 kD.

In a previous paper, the matriptase cDNA clone was transfected into COS-7 cells. Membrane extracts of these cells were compared in an immunoblot to matriptase derived from the conditioned medium of T-47D human breast cancer cells (3). Presumably matriptase is cell-surface bound, similar to expression of MT-SP1 constructs lacking the signal/anchor domain. Under non-reducing conditions, the size of matriptase appears to be the same size as the protein isolated from the breast cancer cells; unfortunately, no molecular weight was designated in the figure, making it difficult to ascertain the size of the proteins. Since the matriptase from the breast cancer cell line was derived from the conditioned media and not from the cell surface, this protein may be cleaved from the surface of the cells or result from an alternatively spliced form of the protein, resulting in a molecular weight that corresponds to a size similar to the predicted matriptase protein (76 kD).

There is strong sequence conservation between MT-SP1 and the mouse homolog epithin (2), which also has a predicted signal anchor domain. If the N-terminal transmembrane domain were untranslated, then divergence would be expected at both the

cDNA and protein level. Instead, strong conservation of amino acids is observed in this N-terminal region, supporting the suggested protein translation and localization suggested for MT-SP1.

The results from PS-SCL implied that the most effective substrate would contain Lys-Arg-Ser-Arg in the P4 to P1 sites respectively. However, since PS-SCL reveals the ideal amino acid for a given position on average, interdependencies between positions are not apparent. However, the clones derived from substrate phage studies (Table Ia) revealed a striking trend: if P3 is basic, then P4 tends to be non-basic (15 of 17 clones); similarly, if P4 is basic, then P3 tends to be non-basic (13 of 15 clones). Thus on average, basic residues are most abundant of the amino acids at P3 and P4, but for each individual sequence, either P3 or P4 is basic, but not both simultaneously. Taking the PS-SCL and the substrate phage together, the ideal sequence should be P4-(Arg/Lys) P3-(X) P2-(Ser) P1-(Arg) P1'-(Ala) and P4-(X) P3-(Arg/Lys) P2-(Ser) P1(Arg) P1'(Ala) where X must be a non-basic amino acid. Although insight into the function of the protease cannot be gained from this substrate specificity alone, the specificity can be used to identify possible macromolecular substrates and these substrates can be tested *in vitro*.

Since MT-SP1 is localized to the extracellular surface of cells, logical substrates might have similar localization and should be cleaved/activated by proteases. This candidate approach revealed that PAR2 and sc-uPA were macromolecular substrates of MT-SP1. PAR2 is highly expressed in human pancreas, kidney, colon, liver and small intestine, and expressed to a lower extent in the prostate, heart, lung and trachea (42). In the small intestine, trypsin may be the physiological activator of PAR2; activation of PAR2 through trypsin cleavage may regulate the epithelium and mediate inflammation and cytoprotection (13). Trypsin may also activate PAR2 in the airways to initiate a bronchoprotective response (14). However, trypsin, most likely, is not the only physiological activator of PAR2, since trypsin is not coexpressed in all tissue types listed above. Another possible activator of PAR2 is mast cell tryptase, which has been shown to activate PAR2 in a tissue culture

system (43). However, the cleavage of PAR2 by tryptase would require the presence of mast cells, which are usually involved in an inflammatory response; thus other activators of PAR2 may exist. MT-SP1 has a similar profile of tissue expression as PAR2 (1), and MT-SP1 and PAR2 are coexpressed in some cell types, including the prostate carcinoma cell line, PC-3 (1, 42). Furthermore, both proteins share the same extracellular surface localization, so it is possible that MT-SP1 may be a natural activator of PAR2.

sc-uPA is the other candidate that was activated by MT-SP1. While the biology of PAR2 still is being elucidated, the biological roles of uPA are well established (see e.g. 17, 18). For example, uPA has been implicated in tumor cell invasion and metastasis; cancer cell invasion into normal tissue can be facilitated by uPA through its activation of plasminogen, which degrades the basement membrane and extracellular matrix. Thus, activators of sc-uPA would be expected to increase invasiveness and possibly the metastatic capacity of tumor cells. The PC-3 prostate cancer cell line coexpresses MT-SP1 (1), uPA (44), and the uPA receptor by immunofluorescence (Takeuchi, T., Shuman, M., and Craik, C. S., unpublished results) allowing cell surface localization of the protease. Thus, MT-SP1 may activate sc-uPA on the surface of PC-3 cells and thereby increase the invasiveness of these cells. Indeed, potent inhibitors of MT-SP1 inhibit the proliferation and metastasis of PC-3 cells in SCID mice (Elfman, F., Takeuchi, T., Conn, M., Craik, C. S., and Shuman, M., unpublished results). However, further studies are being performed to clearly identify MT-SP1 as the selective target of the protease inhibitor. Nevertheless, the finding that MT-SP1 activates sc-uPA may have interesting implications for the role of proteolysis in cancer.

MT-SP1 is a highly active enzyme with  $k_{cat}/K_m$  for synthetic peptide substrate turnover approaching levels of the digestive enzyme trypsin (1). However, MT-SP1 is not a non-specific degradative enzyme. The specificity of MT-SP1 for macromolecular substrates closely matches the specificity determined in PS-SCL and substrate phage display. Thus, while PAR2 is activated, the highly similar PAR1, 3, and 4 are not activated. The substrate sc-uPA is activated, but the plasminogen is not cleaved even

at much higher concentrations of MT-SP1. Interestingly, the activation site of pro-MT-SP1 matches the substrate specificity determined for the active enzyme, and recombinant MT-SP1 was found to autoactivate upon removal of denaturant (1). The high activity of the MT-SP1 catalytic domain may allow residual activity of the pro-enzyme, allowing autoactivation to occur. Thus, MT-SP1 could autoactivate and initiate signaling and proteolytic cascades via activation of PAR2 or sc-uPA. Other membrane-type serine proteases involved in a proteolytic cascades are enteropeptidase (45), which activates trypsinogen in the gut for digestion and hepsin (46), which has been shown to activate factor VIIa in a blood coagulation cascade (47). Other membrane-type serine proteases include TMPRSS2 (48), human airway trypsin-like protease (49) and corin (50). Membrane-type serine proteases as signaling molecules that play key regulatory roles may become more prevalent as these novel proteases are further characterized.

## **Acknowledgments**

We thank Bradley Backes, Francesco Leonetti, and Jonathan Ellman for PS-SCL library synthesis and thank Ibrahim Adiguzel, Yonchu Jenkins, and Sushma Selvarajan for technical assistance and helpful discussions. T.T. was supported by a National Institutes of Health postdoctoral fellowship CA71097 and Department of Defense Prostate Cancer Research Program Postdoctoral Fellowship DAMD17-99-1-9515. This work was supported by National Institutes of Health Grant CA72006, Developmental Research Program of the UCSF Prostate Cancer Center and by the Daiichi Research Center.

## References

1. Takeuchi, T., Shuman, M. A., and Craik, C. S. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 11054-11061
2. Kim, M. G., Chen, C., Lyu, M. S., Cho, E.-G., Park, D., Kozak, C., and Schwartz, R. H. (1999) *Immunogenetics* **49**, 420-428
3. Lin, C.-Y., Anders, J., Johnson, M., Sang, Q. A., and Dickson, R. B. (1999) *J. Biol. Chem.* **274**, 18231-18236
4. Huber, R. & Bode, W. (1978) *Acc. Chem. Res.* **11**, 114-122
5. Pinilla, C., Appel, J. R., Blanc, P., and Houghten, R. A. (1992) *Biotechniques* **13**, 901-905
6. Rano, T. A., Timkey, T., Peterson, E. P., Rotonda, J., Nicholson, D. W., Becker, J. W., Chapman, K. T., Thornberry, N. A. (1997) *Chem. Biol.* **4**, 149-155
7. Thornberry, N. A., Rano, T. A., Peterson, E. P., Rasper, D. M., Timkey, T., Garcia-Calvo, M., Houtzager, V. M., Nordstrom, P. A., Roy, S., Vaillancourt, J. P., Chapman, K., T., and Nicholson, D. W. (1997) *J. Biol. Chem.* **272**, 17907-17911
8. Harris, J. L., Peterson, E. P., Hudig, D., Thornberry, N. A., and Craik, C. S. (1998) Definition and redesign of the extended substrate specificity of Granzyme B. *J. Biol. Chem.* **273**, 27364-27373
9. Matthews, D. J., and Wells, J. A. (1993) *Science* **260**, 1113-1117
10. Matthews, D. J., Goodman, L. J., Gorman, C. M., and Wells, J. A. (1994) *Protein Sci.* **3**, 1197-1205
11. Schechter, I. & Berger, A. (1967) *Biochem. Biophys. Res. Commun.* **27**, 157-162
12. Backes, B.J., Harris, J. L., Leonetti, F., Craik, C. S., and Ellman, J. A., *Nature Biotech.* (2000) **18**, 187-193
13. Kong, W., McConalogue, K., Khitin, L. M., Hollengerg, M. D., Payan, D. G., Bohm, S. K., and Bunnett, N. W. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 8884-8889

14. Steinhoff, M., Vergnolle, N., Young, S.H., Tognetto, M., Amadesi, S., Ennes, S.H., Trevisani, M., Hollenberg, M.D., Wallace, J.L., Caughey, G.H., Mitchell, S.E., Williams, L.M., Geppetti, P., Mayer, E.A., Bunnett, N.W. (2000) *Nature Med.* **6**, 151 - 158
15. Cocks, T. M., Fong, B., Chow, J. M., Anderson, G. P., Frauman, A. G., Goldie, R. G., Henry, P. J., Carr, M. J., Hamilton, J. R., and Moffatt, J. D. (1999) *Nature* **398**, 156-160
16. Miyata, S., Koshikawa, N., Yasumitsu, H., and Miyazaki, K. (2000) *J. Biol. Chem.* **275**, 4592-4598
17. Dano, K. Andreasen, P. A., Grondahl-Hansen, J., Kristensen, P., Nielsen, L. S., & Skriver, L. (1985) *Adv. Cancer Res.* **44**, 139 266
18. Andreasen, P. A., Kjoller, L., Christensen, L., & Duffy, M. J. (1997) *Int. J. Cancer* **72**, 1-22
19. Unal, A., Pray, T. R., Lagunoff, M., Pennington, M. W., Ganem, D., and Craik, C. S. (1997) *J. Virol.* **71**, 7030-7038
20. Weis, K., Dingwall, C. and Lamond, A. I. (1996) *EMBO J.* **15**, 7120-7128
21. Altin, J. G., and Pagler, E. B. (1995) *Anal. Biochem.* **224**, 382-389
22. Backes, B. J. and Ellman, J. A. (1999) *J. Org. Chem.* **64**, 2322-2330
23. Ostresh, J. M., Winkle, J. H., Hamashin, V. T., and Houghten, R. A. (1994) *Biopolymers* **34**, 1681-1689
24. Pereira, P. J., Bergner, A., Macedo-Ribeiro, S., Huber, R., Matschiner, G., Fritz, H., Sommerhoff, C. P., Bode, W. (1998) *Nature* **392**, 306-311
25. McGrath, M. E., Erpel, T., Bystroff, C., and Fletterick, R. J. (1994) *EMBO J.* **13**, 1503-1507
26. Vu, T.-K.H., Hung, D.T., Wheaton, V.I. & Coughlin, S.R. (1991) *Cell* **64**, 1057-1068
27. Nystedt, S., Larsson, A. K., Aberg, H. and Sundelin, J. (1995) *J. Biol. Chem.* **270**, 5950-5955
28. Ishihara, H., Connolly A. J., Zeng D., Kahn M.L., Zheng Y.W., Timmons C., Tram T., Coughlin S. R. (1997) *Nature* **386**, 502-506

29. Kahn, M.L., Zheng Y. W., Huang W., Bigornia V., Zeng D., Moff S., Farese R.V. Jr, Tam C., Coughlin, S.R. (1998) *Nature* **394**, 690-694
30. Ishii, K., Hein, L., Kobilka, B. K., and Coughlin, S. R. (1993) *J. Biol. Chem.* **268**, 9780-9786
31. Schubert, H.-J., Kroell, A., and Leibold, W. (1996) *J. Immunol. Methods* **189**, 89-98
32. Ke, S. -H., Coombs, G. S., Tachias, K., Navre, M., Corey, D. R., and Madison, E. L. (1997) *J. Biol. Chem.* **272**, 16603-16609
33. Coughlin, S. R. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 11023-11027
34. Nielsen, L. S., Hansen. J. G., Skriver, L., Wilson, E. L., Kaltoft, K., Zeuthen, J., and Dano, K. (1982) *Biochemistry* **1982**, 6410-6415
35. Ichinose, A., Fujikawa, K., Suyama, T. (1986) *J. Biol. Chem.* **261**, 3486-3489
36. Kobayashi, H., Schmitt, M., Goretzki, L., Chucholowski, N., Calvete, J., Kramer, M., Gunzler, W. A., Janicke, F., and Graeff, H. (1991) *J. Biol. Chem.* **266**, 5147-5152
37. Goretzki, L., Schmitt, M., Mann, K., Calvete, J., Chucholowski, N., Kramer, M., Gunzler, W. A., Janicke, F., Graeff, H. (1992) *FEBS Lett.* **297**, 112-118
38. Stack, M. S., Johnson, D. A. (1994) *J. Biol. Chem.* **269**, 9416-9419
39. Yoshida, E., Ohmura, S., Sugiki, M., Maruyama, M., Mihara, H. (1995) *Int. J. Cancer* **63**, 863-865
40. Lin, C.I-Y., Wang, J.-K., Torri, J., Dou, L., Sang, Q. A., and Dickson, R. B. (1997) *J. Biol. Chem.* **272**, 9147-9152
41. Shi, Y. E., Torri, J., Yieh, L., Wellstein, A., Lippman, M. E., and Dickson, R. B. (1993) *Cancer Res.* **53**, 1409-1415
42. Bohm, S. K., Kong, W., Bromme, D., Smeekens, S. P., Anderson, D. C., Connolly, A., Kah, M., Nelken, N. A., Coughlin, S. R., Payan, D. G., and Bunnett, N. W. (1996) *Biochem. J.* **314**, 1009-1016
43. Molino, M., Barnathan, E. S., Numerof, R., Clark, J., Dreyer, M., Cumashi, A., Hoxie, J. A., Schechter, N., Woolkalis, M., and Brass, L. F. (1997) *J. Biol. Chem.* **272**, 4043-4049

44. Yoshida, E., Verrusio, E. N., Mihara, H., Oh, D., and Kwaan, H. C. (1994) *Cancer Res.* **54**, 3300-3304
45. Huber, R. & Bode, W. (1978) *Acc. Chem. Res.* **11**, 114-122
46. Leytus, S. P., Loeb, K. R., Hagen, F. S., Kurachi, K., & Davie, E. W. (1988) *Biochemistry* **27**, 1067-1074
47. Kazama, Y., Hamamoto, T., Foster, D. C., Kisiel, W. (1995) *J. Biol. Chem.* **270**, 66-72
48. Poloni-Giacobino, A., Chen, H., Peitsch, M. C., Rossier, C., & Antonarkis, S. E. (1997) *Genomics* **44**, 309-320
49. Yamakoka, K., Masuda, K., Ogawa, H., Takagi, K., Umemoto, N., & Yasuoka, S. (1998) *J. Biol. Chem.* **273**, 11895-11901.
50. Yan, W., Sheng, N., Seto, M., Morser, J., and Wu, Q. (1999) *J. Biol. Chem.* **274**, 14926-14935.
51. Bork, P. & Beckmann, G. (1993) *J. Mol. Biol.* **231**, 539-545
52. Krieger, M., & Herz, J. (1994) *Annu. Rev. Biochem.* **63**, 601-637
53. Perona, J. J. & Craik, C. S. (1997) *J. Biol. Chem.* **272**, 29987-29990

## Figure Legends

Table I: Substrate Specificity of MT-SP1 determined by substrate phage. (a) The unbiased library has P1 fixed as R, while P4, P3, P2, and P1' can encode any amino acid; sequences are shown from round 8. (b) The biased library has P3 fixed as (R/K/Q), P2 fixed as (S/A/G), and P1 fixed as R, while P4, P1', and P2' can encode any amino acid sequences are shown from round 5.

Figure 1: a) 1. The proposed domain structure of human MT-SP1. SA represents a possible signal anchor, CUB represents a repeat first identified in complement components C1r and C1s, the urchin embryonic growth factor and bone morphogenetic protein 1 (51), L represents low-density lipoprotein receptor repeat (52), SP represents a chymotrypsin family serine protease domain (53). The predicted disulfide linkages are shown labeled as C--C. 2. The proposed translation of matriptase. 3. Recombinant, soluble MT-SP1 serine protease domain, where H represents a 6-histidine tag.

b) Immunoblot analysis of recombinant MT-SP1 protease domain and analysis of PC-3 and HeLa cell lysates, *in vitro* transcription/translation of MT-SP1 cDNA, and identification of cell surface biotinylated MT-SP1. Recombinant MT-SP1 is shown in lane 1. MT-SP1 is not expressed by HeLa S3 cells, lane 2. Full length native MT-SP1 appears at 87 kD and the native protease domain 30 kD, lane 3. The protease domain does not appear in immunoblots when the PC3 cell lysates are run under non-reducing conditions [(-)  $\beta$ -mercaptoethanol ( $\beta$ ME)], corroborating the predicted disulfide linkage between the MT-SP1 pro-domain at C604 and the catalytic protease domain at C731 (lane 4). Deglycosylation of the PC-3 cell lysates is shown in lane 6 [(+) PNGase F] compared to similarly treated nondeglycosylated PC3 cell lysates, lane 5. *In vitro* transcription/translation of full-length MT-SP1 cDNA, lane 7. Immunoblotting of cell-surface biotinylated MT-SP1, lane 8.

Figure 2: Immunofluorescence reveals extracellular surface localization of MT-SP1. PC-3 cells treated with MT-SP1 antiserum are shown in (a). HeLa S3 cells treated with MT-SP1 antiserum are shown in (b, negative control).

Figure 3: Activity of MT-SP1 in a P1-Lys Positional Scanning-Synthetic Combinatorial Library. Y-axis is pM of fluorophore released per second. X-axis indicates the amino acid held constant at each position, designated by the one-letter code (n represents norleucine).

Figure 4: Cleavage frequency of amino acids in substrate phage clones. (a) P4-P2 cleavage frequency is shown for the unbiased substrate phage library from rounds 5 and 8. (b) P4-P2 cleavage frequency is shown for the biased substrate phage library from round 5.

Figure 5: The active site from a model of MT-SP1 is shown with the substrate Arg-Phe-Gln-Arg bound in the S1 through S4 subsites respectively. MT-SP1 side chains are shown in red, while the substrate is shown in blue. The protease amino acids are labeled with MT-SP1 numbering and chymotrypsinogen numbering in parentheses.

Figure 6: Activation of PAR2 by soluble MT-SP1 protease domain. A) *Xenopus* oocytes were injected with cRNA encoding the indicated PAR, and protease-triggered  $^{45}\text{Ca}$  release was assessed. Data shown are expressed as fold increase over basal ( $^{45}\text{Ca}$  released in the 10 minutes after agonist addition/ $^{45}\text{Ca}$  released in the 10 minutes before). B) Surface expression of the PARs in (A) was determined by binding of a monoclonal antibody to a FLAG epitope displayed at each receptor's N-terminus. In A and B, the data shown are means of duplicate determinations and the results shown are representative of those obtained in 3 separate experiments.

Figure 7: Activation of sc-uPA with MT-SP1. Data from a representative experiment is shown. 5  $\mu\text{M}$  uPA activation with 1 nM MT-SP1 at 37°C is assayed at specified times by

removing aliquots and monitoring activity at 25°C against the substrate Spectrozyme uK. Activity shown represents a 133-fold dilution from the original reaction mixture. Cleavage of 5  $\mu$ M sc-uPA with 1 nM MT-SP1 at 37°C is examined over time using immunoblot analysis. Sc-uPA is cleaved into an A-chain and B-chain upon activation. Native, active, uPA, which is used as a control (+), has different glycosylation compared with recombinant sc-uPA used in the assay, leading to differences in molecular weight. Unreacted sc-uPA incubated under the same conditions is shown in (-).

Table I

a) Unbiased library clone number	P4	P3	P2	P1	P1'
1	V	T	G	R	S
2	V	R	G	R	S
3	A	Q	G	R	M
4	R	E	G	R	M
5	R	E	G	R	T
6	G	S	G	R	W
7	-	Q	G	R	R
8	G	Q	G	R	-
9	K	Q	G	R	A
10	R	K	G	R	S
11	G	R	G	R	-
12	G	K	G	R	T
13	E	R	G	R	S
14	A	R	G	R	R
15	K	M	G	R	R
16	R	R	G	R	T
17	P	L	G	R	S
18	K	E	G	R	L
19	R	E	G	R	V
20	R	M	G	R	A

b) Biased library clone number	P4	P3	P2	P1	P1'	P2'
1	L	K	S	R	V	K
2	S	K	S	R	T	L
3	F	Q	C	R	V	F
4	L	K	S	R	L	S
5	S	K	S	R	L	S
6	F	K	A	R	N	C
7	H	K	G	R	A	K
8	F	Q	S	R	M	E
9	I	R	S	R	Y	V
10	Y	K	S	R	N	L
11	W	K	S	R	S	N
12	V	K	S	R	T	S
13	V	N	C	R	T	N
14	S	K	A	R	T	T
15	L	K	S	R	V	H
16	A	Q	S	R	M	S
17	I	K	G	R	M	A
18	D	Q	S	R	M	T
19	R	Q	S	R	L	C
20	F	Q	S	R	G	N
21	V	K	S	R	L	C

Figure 1a

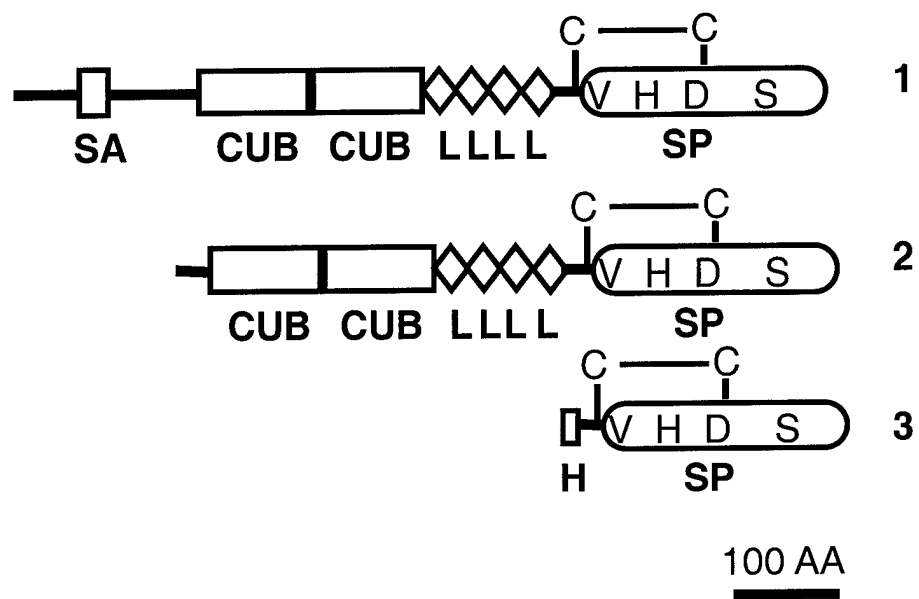


Figure 1b

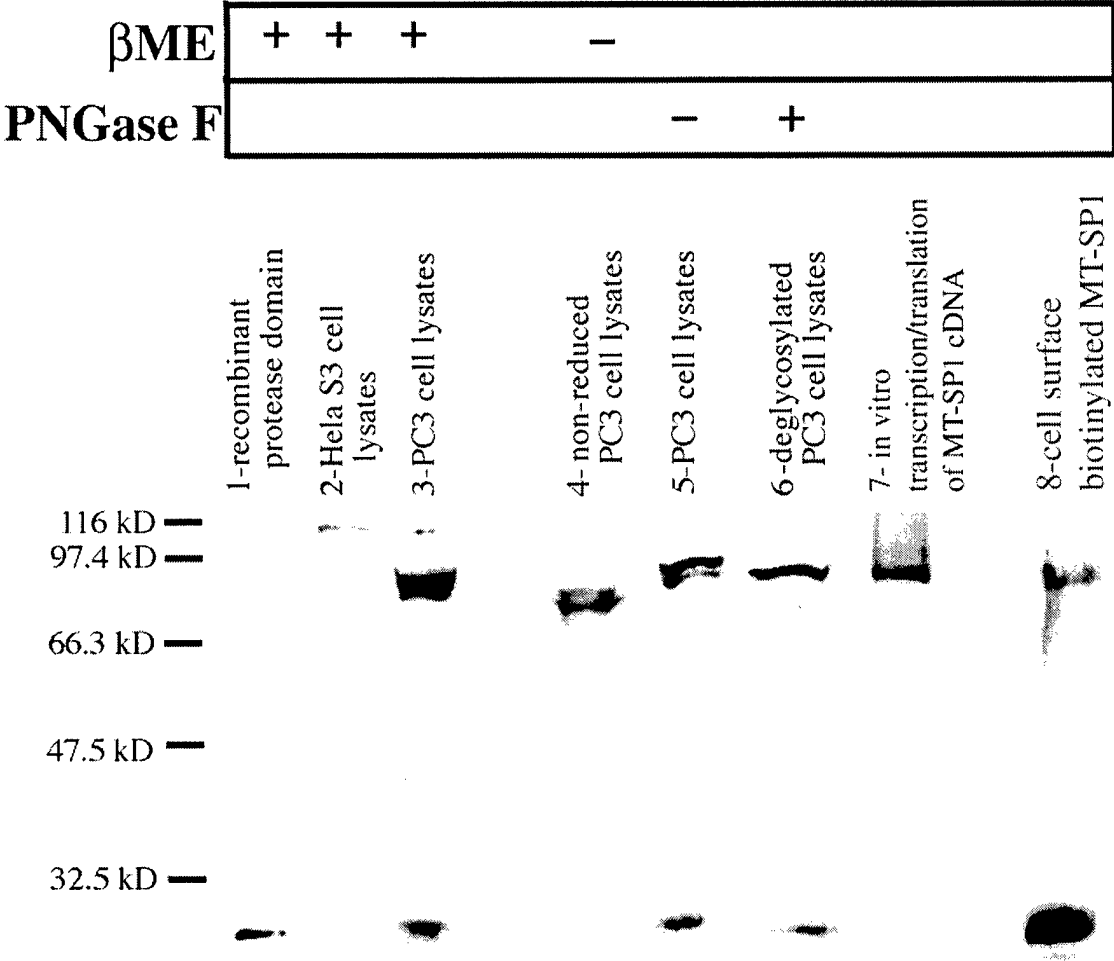
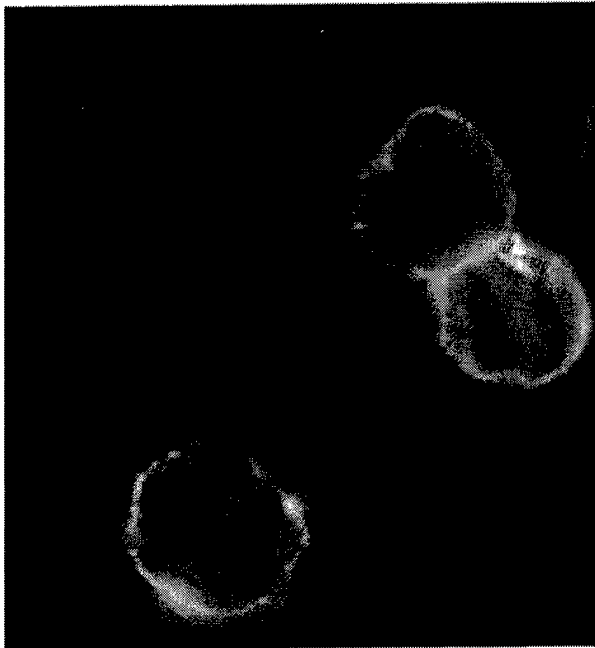


Figure 2

a)



b)

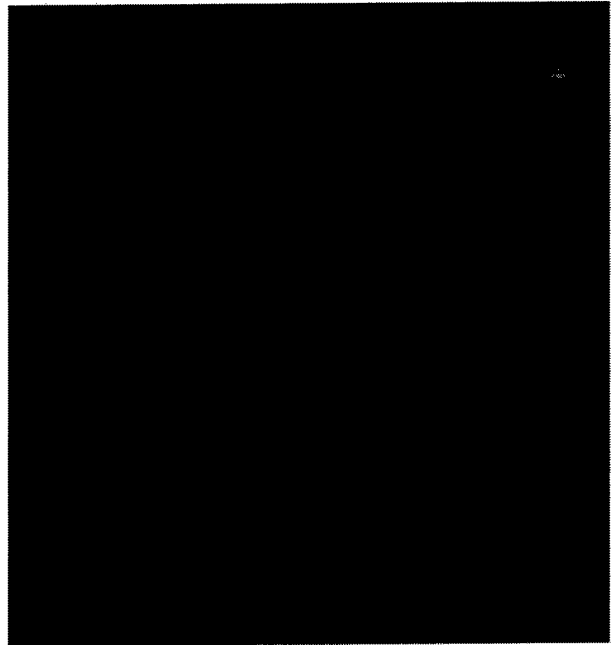


Figure 3

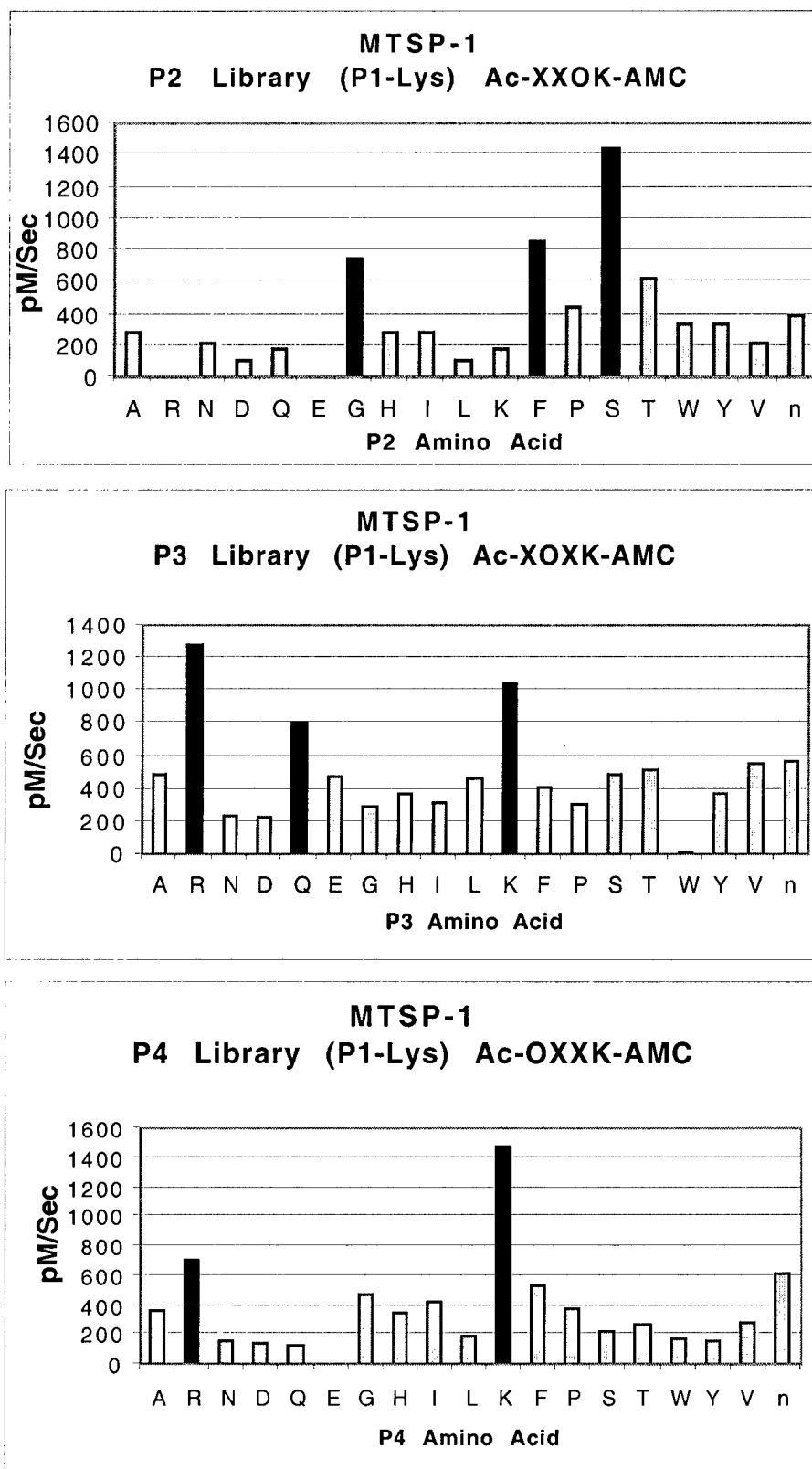


Figure 4a

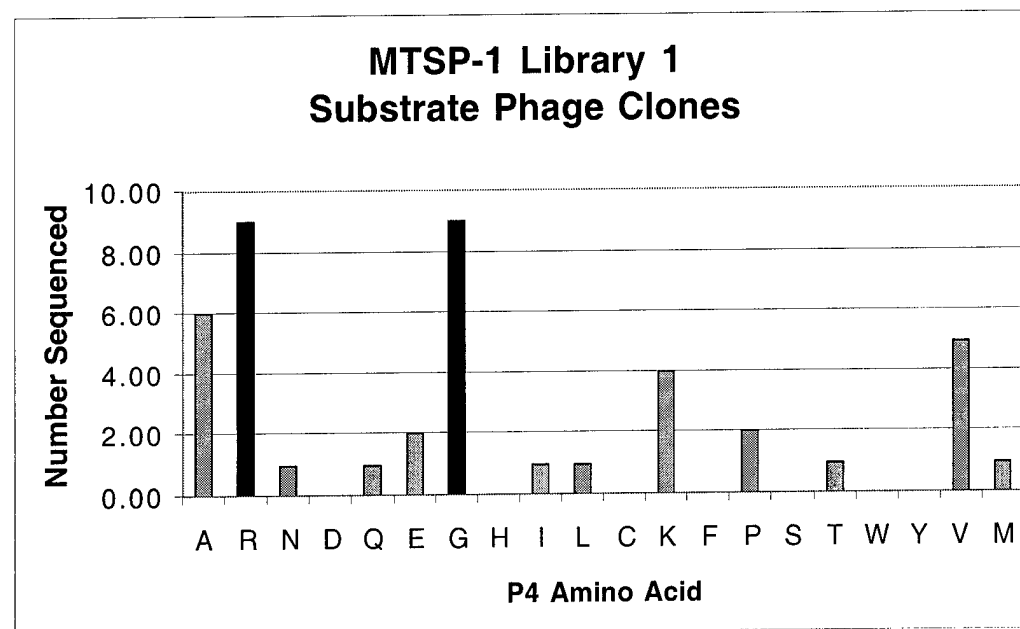
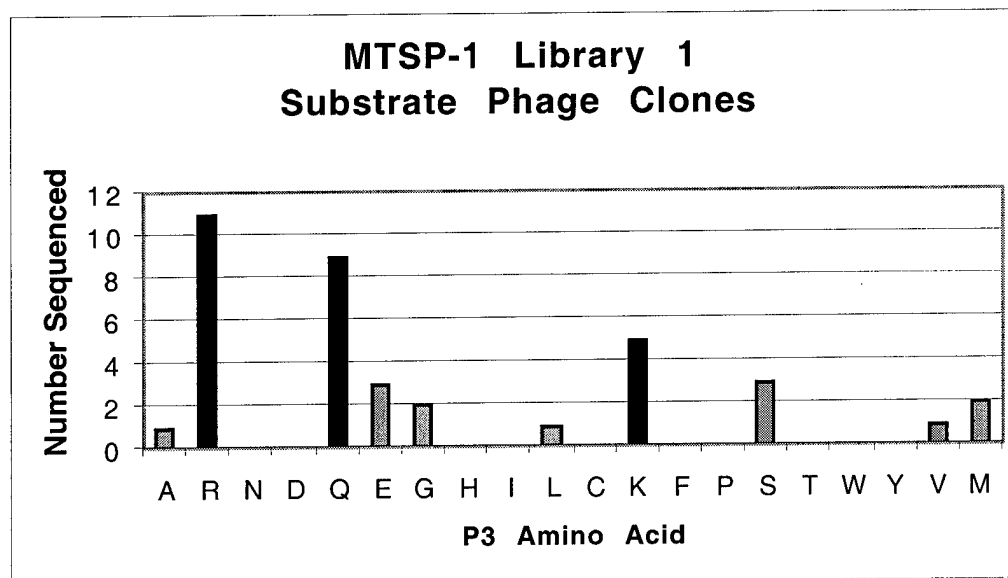
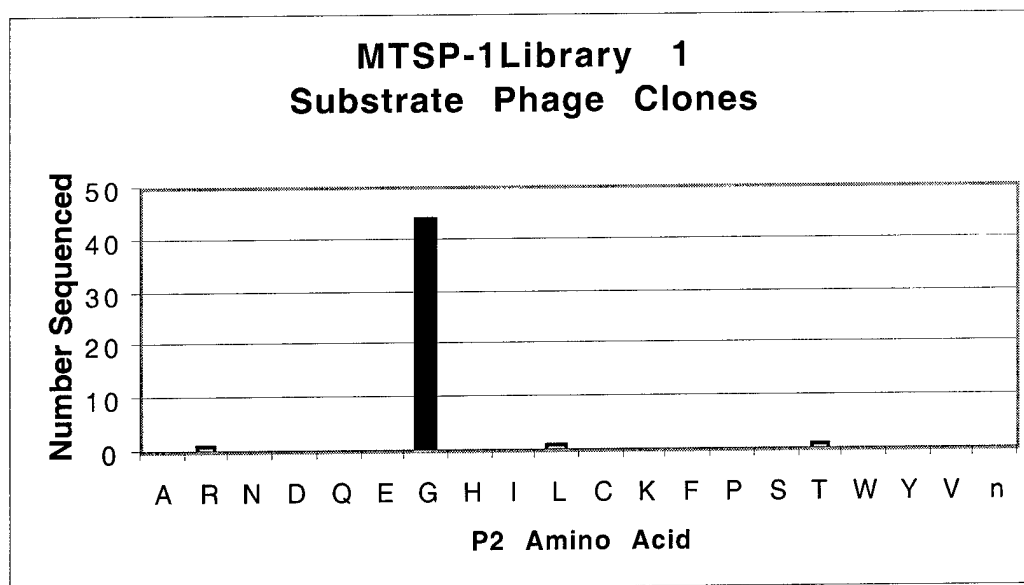


Figure 4b

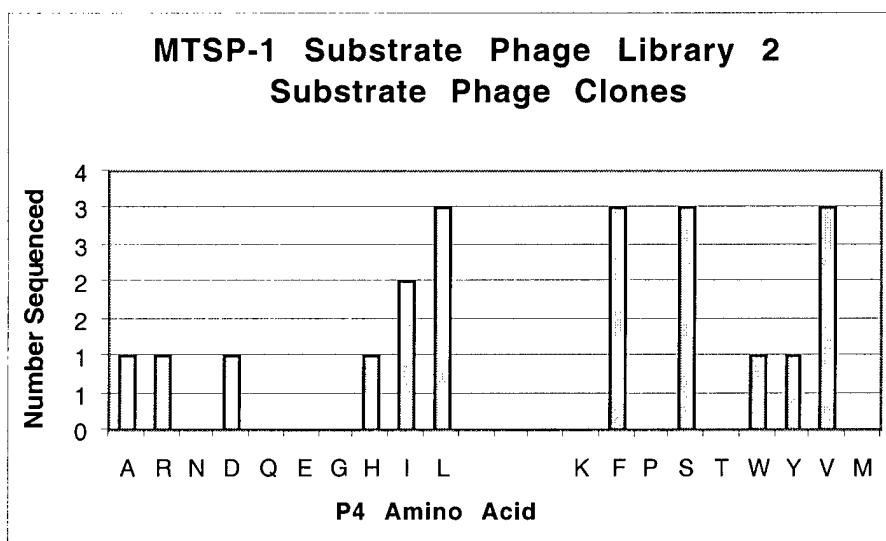
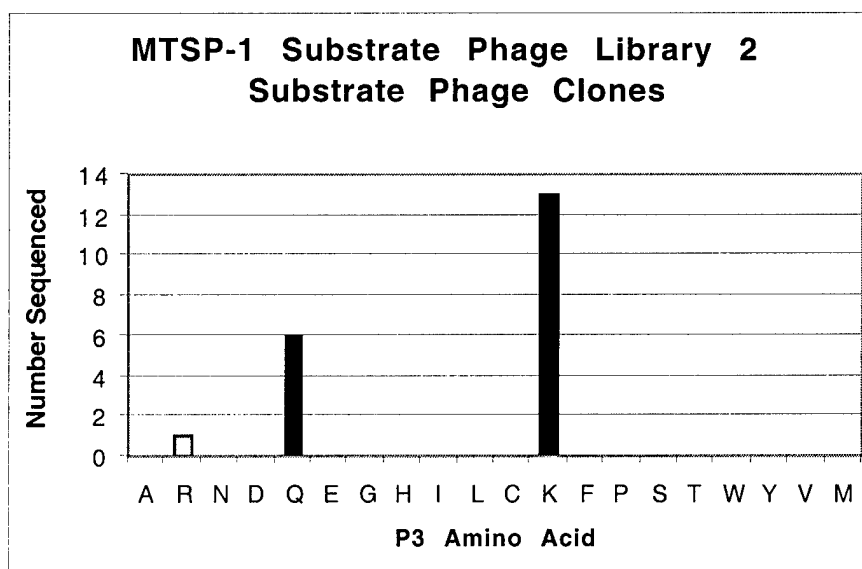
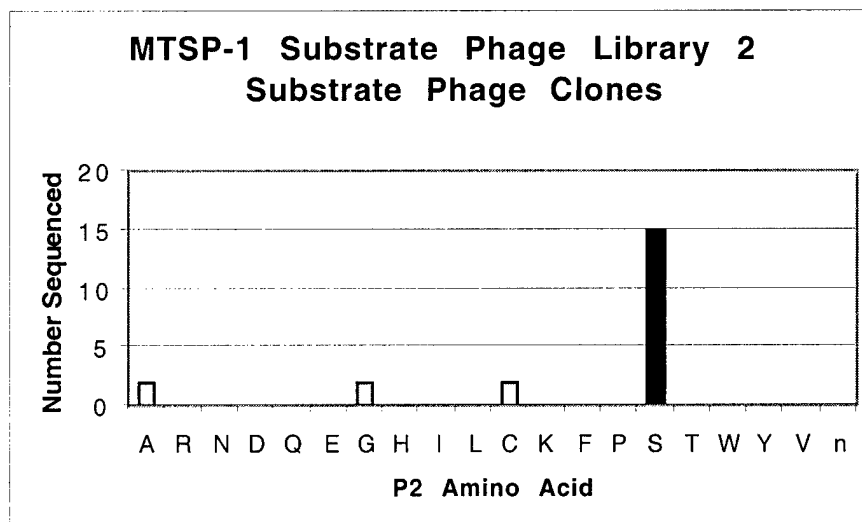


Figure 5

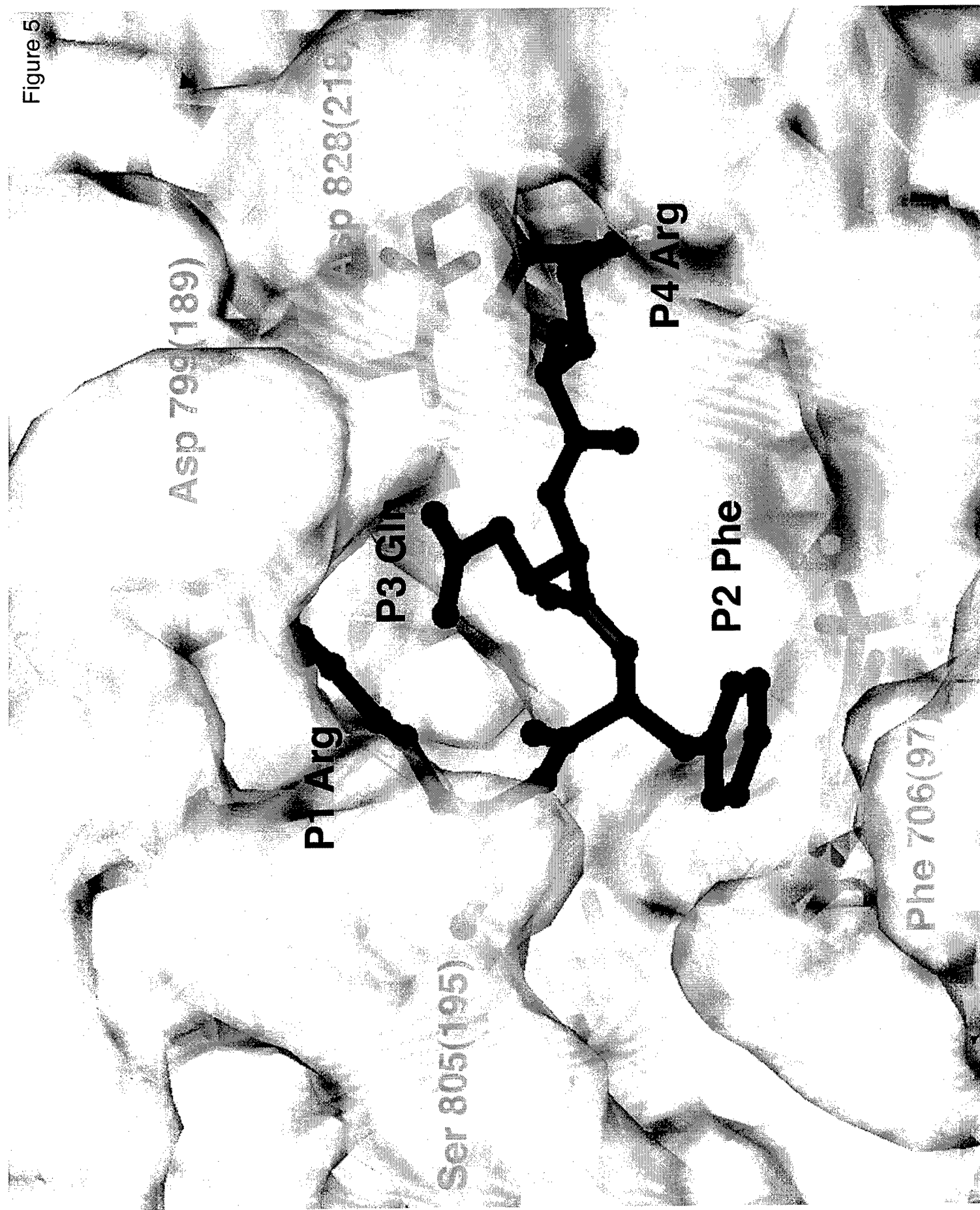
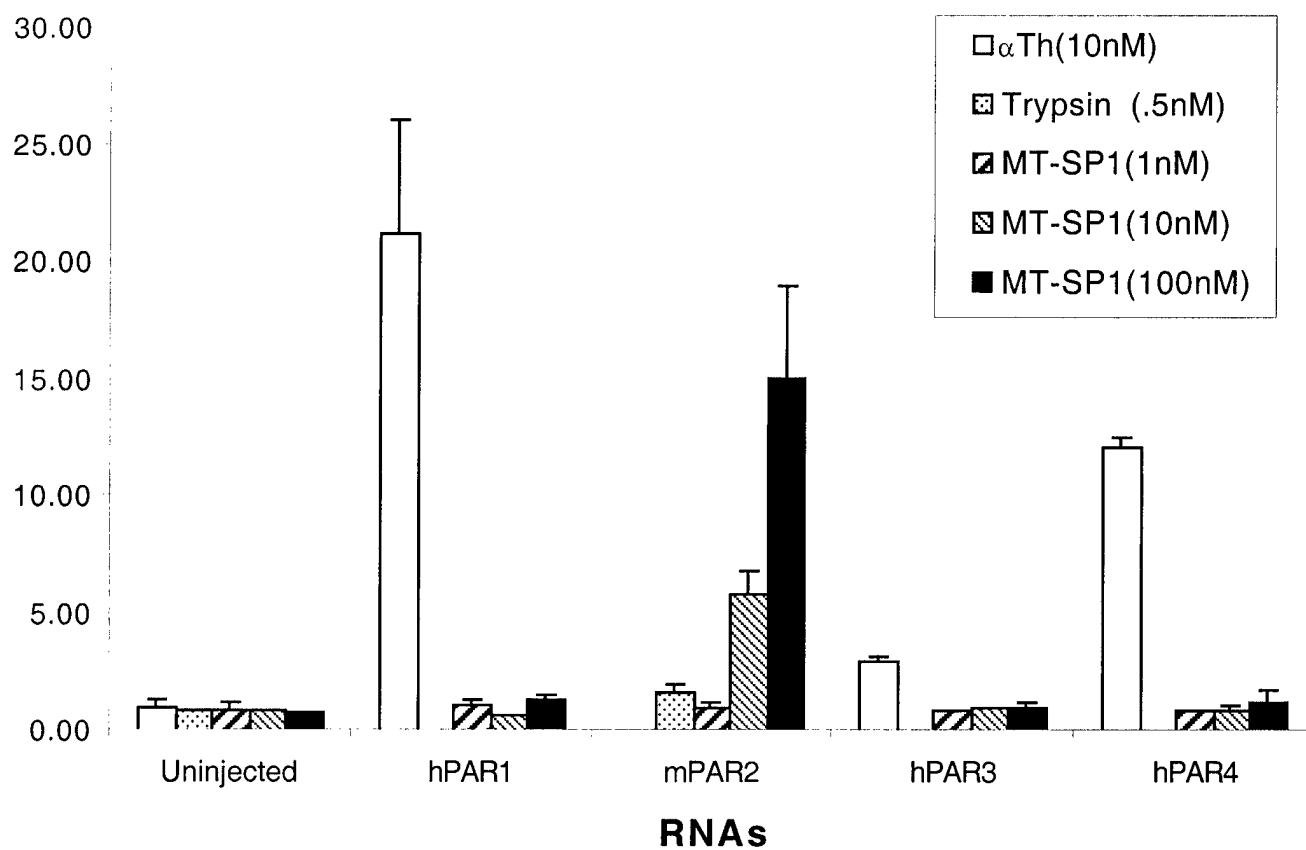


Figure 6

a)

### MT-SP1 Activation of PARs



b)

### PARs Expression

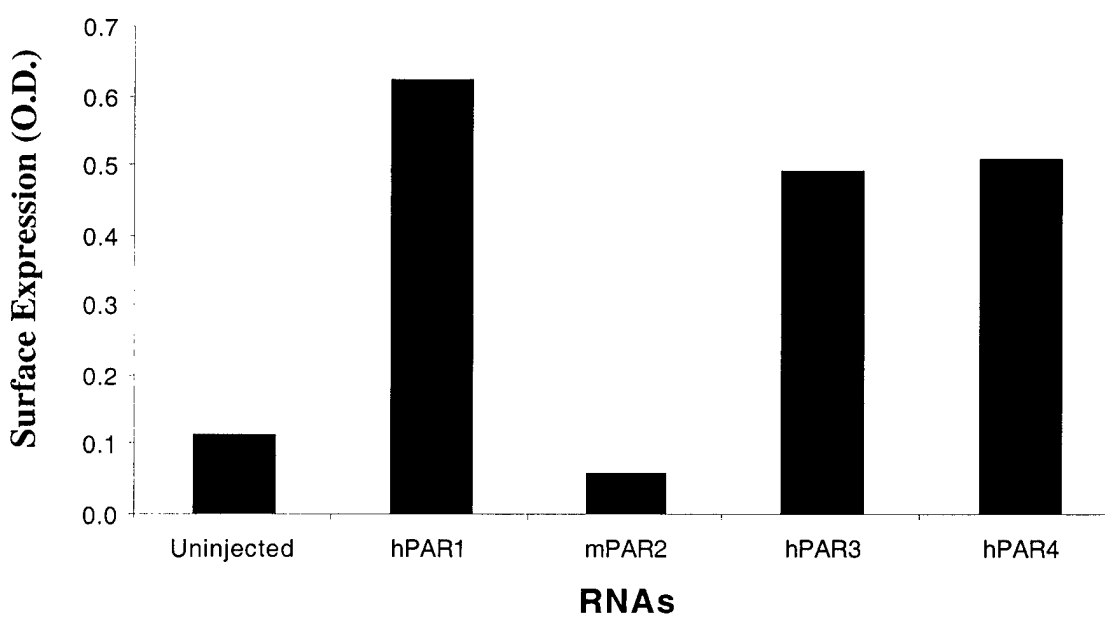


Figure 7

